AD_____

Award Number: DAMD17-01-1-0096

TITLE: The Role of Human Spectrin SH3 Domain Binding Protein 1 (HSSH3BP1) in Prostatic Adenocarcinoma

PRINCIPAL INVESTIGATOR: Leszek Kotula, M.D.
Jiliu Xu, M.D.
Jill A. Macoska, Ph.D.
Piotr Kozlowski, M.D., Ph.D.
Magdalena Martinka, M.D.
Edward C. Jones, M.D.

CONTRACTING ORGANIZATION: Research Foundation for Mental Hygiene, Incorporated
Albany, New York 12229

REPORT DATE: September 2003

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 2003	Annual (20 Aug 02-19 Aug 03)

4. TITLE AND SUBTITLE
The Role of Human Spectrin SH3 Domain Binding Protein 1 (HSSH3BP1) in Prostatic Adenocarcinoma

5. FUNDING NUMBERS
DAMD17-01-1-0096

6. AUTHOR(S)
Leszek Kotula, M.D., Jiliu Xu, M.D., Jill A. Macoska, Ph.D., Piotr Kozlowski, M.D., Ph.D., Magdalena Martinka, M.D., Edward C. Jones, M.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Research Foundation for Mental Hygiene, Incorporated
Albany, New York 12229

E-Mail: kotulal@rcn.com

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates. All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

Prostate cancer is one of the leading causes of cancer-related deaths in the United States and a leading diagnosed cancer in American men. Newly diagnosed cases of prostate cancer approach rapidly the number of 200, 000 cases per year. Genetic alterations of tumor suppressor genes are one of the most common causes of prostate cancer tumorigenesis. Our group identified Hssh3bp1 as a candidate prostate tumor suppressor gene. In this research we are testing the tumor suppressor function of the candidate gene in prostatic adenocarcinoma using *in vitro* and *in vivo* assays. Using developed cell lines expressing Hssh3bp1 we established a candidate region of Hssh3bp1 responsible for growth regulation. Within this region we mapped a tyrosine phosphorylation site by c-Abl tyrosine kinase. Another set of experiments established that expression of Hssh3bp1 up-regulated expression levels of c-Abl tyrosine kinase in prostate cell lines. Further work is under way to develop a detail understanding what is the mechanism of growth regulation by Hssh3bp1 and Abl kinase. It is hoped that this work will help us to understand prostate cancer tumorigenesis and ultimately lead to better diagnosis and therapy strategies.

14. SUBJECT TERMS
Gene expression, tumor biology, cell growth, tumor suppression, Tumorigenesis, molecular genetics, protein interactions

15. NUMBER OF PAGES
14

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction

Prostate cancer is one of the leading causes of cancer-related deaths in the United States (over 41, 000 per year) and a leading diagnosed cancer in American men (43% of all diagnosed cancer in men). Newly diagnosed cases of prostate cancer approach rapidly the number of 200, 000 cases per year. Genetic alterations of tumor suppressor genes are one of the most common causes of neoplastic transformation leading to tumorigenesis including prostate cancer tumorigenesis. Inactivation of one or more tumor suppressor genes is thought to be the most common cause of prostatic adenocarcinoma. Our group identified such candidate tumor suppressor gene. The gene was originally named Hssh3bp1 for its binding properties to spectrin SH3 domain (*human spectrin SH3 domain binding protein 1*).

In this research we propose to test the tumor suppressor function of a candidate gene in prostatic adenocarcinoma using *in vitro* and *in vivo* assays. The work is directed at understanding what is the mechanism of loss of hssh3bp1 expression in prostatic cells lines and tumors, and will test potential tumor suppressive role of hssh3bp1 in nude mice. Hssh3bp1 is a potential regulator of macropinocytosis. Macropinocytosis can be upregulated by growth factors, which in turn promote tumor growth; we propose that Hssh3bp1 is a negative regulator of macropinocytosis and cell growth. To learn more about possible mechanisms of Hssh3bp1 tumor suppressor function we will determine whether Hssh3bp1 mutations affect macropinocytosis of prostate cells and determine molecular events underlying this effect. Although it is possible that Hssh3bp1 is not involved in biogenesis of prostate cancer, after completion of the proposed work we will know more about the function of the protein in human prostate. On the other hand, with the identification of hssh3bp1 as a tumor suppressor gene in prostate cancer, it is likely to lead to subsequent hypotheses and research on the hssh3bp1 role in prostate tumorigenesis. This, in turn, is likely to lead to a better diagnosis, treatment, and possibly prevention of this deadly disease.

Body

The following are the aims of proposal as defined in the original Statement of Work:

Aim 1. To determine whether Hsshb3p1 is not expressed in some prostate tumors due to presence of mutations.

a. Search for mutations of the hssh3bp1 cDNA and gene in prostate tumor cell lines and primary prostate tumors (30 cases).
b. Determine pattern of hssh3bp1 expression in primary prostate tumors. Correlate the pattern of hssh3bp1 expression with the tumor grade and stage (100 cases).

Aim 2. To determine whether the Hssh3bp1 gene carries a tumor suppressor function *in vivo*.

a. Evaluate tumorigenicity of prostate cell lines containing mutated hsshb3p1 in athymic nude mice and in soft agar assay. Evaluate the tumorigenecity of cell lines transfected with the hssh3bp1 antisense plasmids in athymic nude mice and in soft agar assay.
b. Identify a region in hssh3bp1 responsible for the tumor suppression function.

Aim 3. To identify a potential mechanism and a signal transduction pathway involved in the tumor suppression function of Hssh3bp1.

a. Determine the role of hssh3bp1 mutations in macropinocytosis of prostate cell lines.
b. Determine the role of growth factors, PI3-kinase, and the 200-kDa spectrin-like protein in the function of hssh3bp1.

We have initiated the work towards all three Aims of the grant application.

Progress towards Aim 1.

The rationale for these experiments is that if Hsshb3p1 carries tumor suppressor function the gene mutations inactivating its function must exist in primary tumors and in tumor cell lines. Although loss of the hssh3bp1 expression may be due to other possibilities including downregulation of a signal transduction pathway(s) involving the gene in the prostate we will specifically search for mutations of Hsshb3p1 because this suggests a tumor suppression function.

We are in the process of collection of prostate tissue biopsies from the local hospital (St. Vincent's Hospital, Staten Island, NY). We have collected 16 specimen up to date positive for prostatic adenocarcinoma. We are in the process of sequencing of Hssh3bp1 cDNA from these specimens (Aim 1a).

Progress towards Aim 1b: We have ordered the prostate tissue array from Department of Pathology, Yale University, New Haven, CT. In total 100 tumor cases with case-matched non-tumor controls will be evaluated for expression of Hssh3bp1 by immunochemistry. We feel that availability of prostate tissue arrays may provide us with much better standardized tissue material (i.e. all tumor cases with non-tumor controls are on the

same slide) than studies of tissue from various sources that we originally proposed. The goals of this Aim remain unchanged.

Progress towards Aim 2.

The rationale for these experiments is to test the hypothetical tumor suppressor function of Hssh3bp1 by complementation assays (Aim 2a). We hypothesized that tumorigenicity of some prostate cell lines is due to inactivation of Hssh3bp1 function. We determined that LnCaP cell lines, ATCC CRL-10995 and -1740 contain an exon-skipping Hssh3bp1 mutation (Macoska *et al.*, 2001). Thus it is possible that Hssh3bp1 function is impared in these cell lines. The goal of the complementation experiments is to transfect a correct copy of Hssh3bp1 gene to cells, restore the gene expression, and examine whether this will suppress malignant phenotype of tumorigenic cells. The experimental plan included establishment of stable clones, testing their growth characteristic by growth assay and colony formation in soft agar (*in vitro* assays) as well as testing of their malignant phenotype by tumorigenicity studies in nude mice (*in vivo* assay).

Determination of candidate growth/tumor suppression pathways involving Hssh3bp1.

Last year we reported establishement of prostate cancer cell lines stably transfected with Hssh3bp1. Two such clones of LnCaP expressing Hssh3bp1, NG18-1 and NG18-10, showed significant reduction of growth (about one third) in comparison to the mock control (NG3-1). In fact the selected cell lines grew so slowly in soft agar that we were not able to perform colony formation assay in soft agar even upon several attempts and modification of the method (data not shown).

We decided to use developed cell lines as model systems to determine potential mechanisms of Hssh3bp1 growth control function. We used microarray expression analysis to determine pattern of gene expression in these cell lines. Duplicate samples of total RNA from Hssh3bp1 expressing LnCaP cell lines, NG18-1 and NG18-10, and mock cell line N3G-1 were hybridized to gene chips containing approximately 26, 000 genes. The analysis was performed using Affymetrix Human Array Expression U133A GeneChips at the Affymetrix Microarray Core, The University of Michigan School of Medicine, Ann Arbor, Michigan. The complete analysis (see attached CD in the Appendix) contains a lot of information that will be extremely useful in establishing the role of Hssh3bp1 in cell growth. In general, stable expression of Hssh3bp1 in LnCaP cell affected expression of several groups of proteins (due to very large amount of information only few examples are given): cytoskeletal proteins (such as adducin, ARP2, alpha 3 tubulin), growth factors receptors (EGFR, human insulin-like growth factor receptor), cell cycle regulators (cyclins G2, D3, and E2), proteins associated with endocytosis (RAB5, ADP ribosylation factor, vesicle docking protein VDP), proteins regulating apoptosis (such as several Bcl2-interacting proteins, apoptosis inhibitor 5), and proteins regulating t-RNA metabolism and processing. Of special interest is of course change of pattern of some prostate-specific genes such as kallikrein 2 known as also PSA, androgen receptor, and prostate differentiation factor. All data (including statistical analysis) is provided on the attached CD. These data will allow us to form specific testable hypotheses to understand Hssh3bp1 role in growth of prostate cells and tissue. We plan to use the data and apply for subsequent grant support from the US Army Department of Medical Research and Materiel Command.

Tumorigenicity assay using was not attempted with the LnCaP-Hssh3bp1 cell lines due the fact that colony assay did not work. Based on consultation with our collaborator on this grant, Dr. Jill Macoska (University of Michigan, Ann Arbor , MI), it would be unlikely that the transfected cells survived in the animals.

Establishement of a candidate Hssh3bp1 region responsible for growth control *in vitro.*

Our earlier findings indicate that two LnCaP cell lines lack expression of normal copy of isoform 2 of Hssh3bp1 (Macoska et al., 2001). Both cell lines contain the Hssh3bp1 exon 6-skipping mutation that result in expression of aberrant RNA transcript and lack of normal isoform 2 expression on the protein level.

Restoration of isoform 2 expression, with exon 6 sequences present, in LnCaP cell resulted in growth inhibition. This strongly suggest that exon 6 contains sequence which is critical for growth control function of Hssh3bp1. In fact, we determined that this region is absolutely necessary for binding to c-Abl tyrosine kinase SH3 domain (data not shown). Exon 6 sequences also contains the major tyrosine phosphorylation site of Hssh3bp1 (see Progress towards Aim 3 below).

Progress towards Aim 3

The major goal of this aim is to identify potential signal transduction mechanism(s) involving Hssh3bp1. In Aim 3b we hypothesized that phosphorylation of Hssh3bp1 occurs following various treatments of cells. It was not known whether and by what enzyme Hssh3bp1 is phosphorylated. However, we hypothesized that Abl kinase is a candidate enzyme since it binds to Hssh3bp1 (Ziemnicka-Kotula *et al.*, 1998). We established that Hssh3bp1 is phosphorylated by Abl kinase *in vitro*. Now, we mapped the tyrosine phosphorylation site of Hssh3bp1 within the exon 6-encoded sequence, and determined that Hssh3bp1 expression upregulates c-Abl tyrosine kinase levels in LnCaP cell lines. These studies identify c-Abl tyrosine kinase as a major regulator of Hssh3bp1 function in LnCaP prostate cells. We are preparing the manuscript describing these studies.

Tyrosine 213 is the major phosphorylation site of Hssh3bp1 by Abl kinase *in vitro* (Fig. 1).

Exon 6-encoded sequence of Hssh3bp1 contains two tyrosine residues, Y198 and Y213. We hypothesized that these residues are phosphorylated by c-Abl tyrosine kinase. We established that in the N-terminal half of Hssh3bp1 these two tyrosines are the only candidate phosphorylation sites of c-Abl kinase *in vitro* (the mutant lacking exon 6, and containing the N-terminal half of Hssh3bp1 was not phosphorylated, data not shown). Subsequent mutagenesis and phosphorylation experiments established that tyrosine 213 is the major phosphorylation site of Hssh3bp1 in the N-terminal region of the protein. Goal of next set of experiments is to determine whether this tyrosine is also phosphorylated in cultured cells by c-Abl kinase.

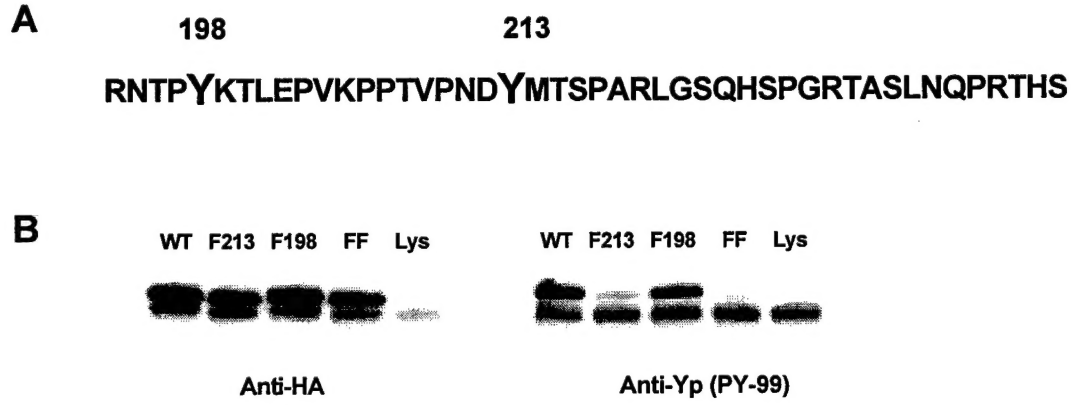


Figure 1. *Phosphorylation of tyrosine 213 by c-Abl tyrosine kinase.* ***A,*** Amino acid sequence of Hssh3bp1 exon 6. Tyrosine residues are depicted by larger font and numbered above the sequence according to position in Hssh3bp1 sequence (Ziemnicka-Kotula *et al.*, 1998). ***B,*** Hssh3bp1 polypeptides containing the N-terminal half of Hssh3bp1 and indicated mutations of tyrosine residues were subjected to *in vitro* kinase reaction with c-Abl tyrosine kinase. Polypeptides were separated on SDS-Tricine polyacrylamide gels (7%) followed by blotting onto the PVDF membrane. Left panel represents the membrane blotted with anti-HA antibody (HA epitope was introduced at the C-terminus of each polypeptide). Right panel represents the same polypeptides blotted with anti-phosphotyrosine antibody PY-99. ***WT***, wild type polypeptide; ***F213***, polypeptide containing fenyloalanine replacement of tyrosine 213; ***F198***, the polypeptide containing

fenyloalanine replacement of tyrosine 198; ***FF***, the polypeptide containing fenyloalanine replacements of tyrosine 198 and tyrosine 213*;* ***Lys***, lysate with no Hssh3bp1 cDNA. Note, the polypeptide F198 containing intact tyrosine 213 shows similar phosphorylation levels to the wild type polypeptide (compare lanes WT and F198 in the right panel).

Expression of Hssh3bp1 upregulates c-Abl kinase in LnCaP cells.

The fact that Hssh3bp1 is a substrate of and a binding partner of c-Abl kinase suggested to us that these two proteins are closely regulated. Therefore we examined levels of c-Abl kinase in stable cell lines of LnCaP transfected with Hssh3bp1. In fact, expression of Hssh3bp1 isoform 2 in LnCaP cells increased levels of c-Abl kinase (Fig. 2) and this correlates with growth inhibition in these cell lines (growth assay data were demonstrated in the last report). c-Abl kinase is known to regulate apoptosis therefore Hssh3bp1 regulation of c-Abl kinase levels could be one of the mechanism by which Hssh3bp1 regulates growth of LnCaP cells. We are pursuing experiments testing this hypothesis right now. These data will be part of a manuscript under preparation.

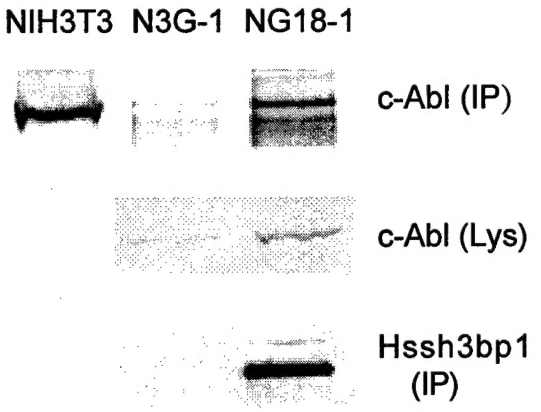


Figure 2. *Expression of recombinant Hsshb3p1 isoform 2 increases expression of c-Abl tyrosine kinase in LnCaP cells.* Total cell lysates (***Lys***), or immunoprecipitates (***IP***) were separated on SDS-Tricine polyacrylamide gels (7%) followed by blotting onto the PVDF membrane. c-Abl was immunoprecipitated with polyclonal antibody K12 (Santa Cruz Biotechnology) and blotted with mAb 8E9, Hssh3bp1 was immunopreciptated with the mAb 4E2 and blotted with polyclonal antibody Ab-2 (Ziemnicka-Kotula *et al.*, 1998). ***N3G-1*** and ***NG18-1*** represent a mock control and a clone stably transfected with Hssh3bp1 isoform 2, respectively. Control immunoprecipation of c-Abl was performed from NIH 3T3 cells (***3T3***). Note increased intensity of c-Abl in NG18-1 clone expressing Hssh3bp1.

Key Research Accomplishements

- **Determination of candidate growth/tumor suppression pathways involving Hssh3bp1.**
- **Establishement of a candidate Hssh3bp1 region responsible for growth control *in vitro*.**
- **Tyrosine 213 is the major phosphorylation site of Hssh3bp1 by Abl kinase *in vitro*.**
- **Expression of Hssh3bp1 regulates c-Abl kinase levels in LnCaP cells.**

Reportable Outcomes

Development of stable prostate cell lines expressing Hssh3bp1.

The LnCap and PC3 cell lines expressing Hssh3bp1 established in our laboratory will be available to scientific community upon publication of the results of this work.

W obtained the grant from NIH based on some of results of this work. The grant is entitled "Regulation of Macropinocytosis by Hssh3bp1"(R01 NS 044968-01A1). The above grant received score 160 and 3.4 percentile.

Microarray expression data (see attached CD disk) will be available to scientific community upon publication of the results of this work.

Conclusions

Three major conclusion of the presented progress of work are:

1. Exon 6 is a candidate Hssh3bp1 region responsible for growth control *in vitro*.

2. Tyrosine 213 is the major phosphorylation site of Hssh3bp1 by Abl kinase *in vitro*.

3. Expression of Hssh3bp1 regulates c-Abl kinase levels in LnCaP cells.

This year we moved further towards establishing Hssh3bp1 as a major regulator of c-Abl kinase function. All data support that point. Exon 6 of Hssh3bp1 contain sequences that are critical for binding to Abl SH3 domain and growth assay data suggest that this is the region most important for Hssh3bp1 growth regulation. Expression of Hssh3bp1 in LNCaP cells inhibits growth of cells (these results were reported last year) and up-regulates c-Abl kinase expression. Identification of the major phosphorylation site of Hssh3bp1 by c-Abl tyrosine kinase provides both starting point for protein biochemical studies of Hssh3bp1 in prostate tumors as well as a specific change (or marker) that can be compared in tumor vs. normal tissue.

References

1. Macoska, J. A., Xu, J., Ziemnicka, D., Schwab, T., Rubin, M., and L. Kotula. 2001. Loss of expression of human spectrin Src homology domain binding protein 1 is associated with 10p loss in human prostatic adenocarcinoma. *Neoplasia* 3: 99-104.

2. Schwab , T.S., Steward, T., Lehr, J., Pienta, K.J., Rhim, J.S., and J. A. Macoska. 2000. Phenotypic characterization of immortilized Normal and primary tumor-derived human prostate epithelial cell cultures. *The Prostate* 44: 164-171.

3. Xu, J., Ziemnicka, D., Merz, G.S., and L. Kotula. 2000. Human spectrin SH3 domain binding protein 1 regulates macropinocytosis in NIH 3T3 cells. *J Cell Sci.* 113:3805-3814.

4. Xu, J., Ziemnicka, D., Scalia, J., and L. Kotula. 2001. Monoclonal antibodies to α spectrin Src homology domain 3 associate with macropinocytic vesicles in nonerythroid cells. *Brain Res.* 898:171-177.

5. Ziemnicka-Kotula, D., Xu, J., Gu, H., Potempska, A., Kim, K. S., Jenkins, E. C., Trenkner E., and L. Kotula. 1998. Identification of A Candidate Spectrin SH3 Binding Protein Suggests A General Mechanism of Association of Tyrosine Kinases with the Spectrin-Based Membrane Skeleton. *J. Biol. Chem.* 273: 13681-13692.

Appendix

1 CD containing microarray expression data and initial statistical analysis of data (this file is also included on the disk).

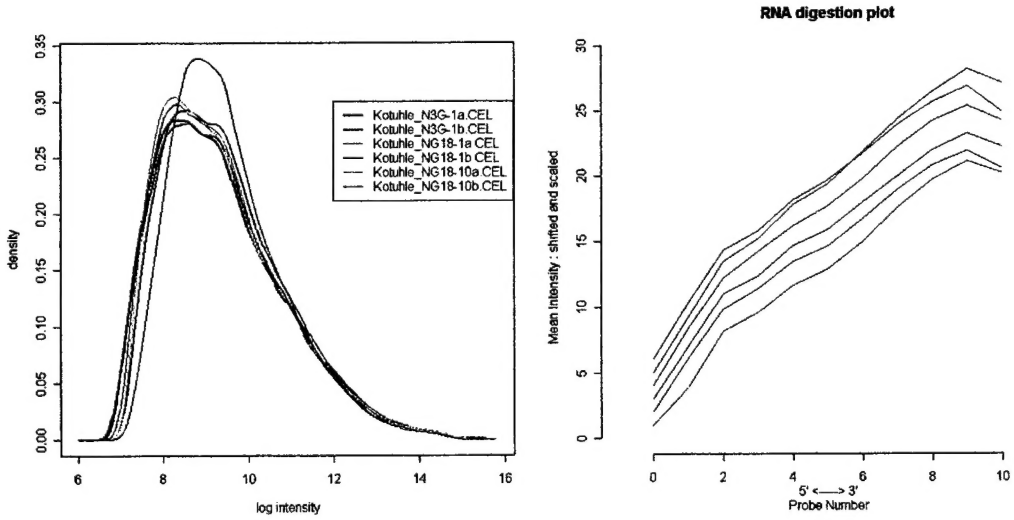
Statistical Analysis of Affymetrix GeneChips for Kotula Lab 6/18/2003

The following analysis is based on six HG-U133A Affymetrix GeneChips that were processed in the UMCCC Microarray Core in early April 2003. The chip names and sample types are as follows:

Filename	Sample Type
Kotula_N3G-1a.CEL	Empty Vector
Kotula_N3G-1b.CEL	Empty Vector
Kotula_NG18-1a.CEL	Transfected Isolate 1
Kotula_NG18-1b.CEL	Transfected Isolate 1
Kotula_NG18-10a.CEL	Transfected Isolate 10
Kotula_NG18-10b.CEL	Transfected Isolate 10

I computed expression values using a robust multi-array average (RMA), which is implemented in the affy library of the Bioconductor package of the statistical language R. This method first normalizes all the perfect match (PM) probe data to have the same distribution using a quantile normalization procedure. Next, it fits a model to the data that accounts for both probe specific intensity and chip specific intensity. The probe specific intensity accounts for differences in the intensity values for the probes specific to a given gene. Since these probes all measure the same mRNA concentration, any differences are clearly not biological in nature. The chip specific intensity is the average intensity of the probes after accounting for the probe specific intensity. This is the expression value reported by the software.[1]

Prior to computing the expression values, I check the quality of the data. For instance, the quantile normalization assumes that the distribution of PM probes is the same on each chip, only differing by location or scale. I check this assumption using a density plot.



The density plot indicates that the distribution of PM probe data on each chip is very similar. I also made a RNA digestion plot, which gives an idea of the extent of RNA

degradation and/or the extent of 3′ bias in the first strand synthesis. This is done by ordering the probes from 5′ to 3′ and then taking the average of each ordered probe. Since degradation occurs in the 5′ to 3′ direction, any slope to these lines indicates that either some degradation has occurred or that the first strand synthesis did not go to completion. I would prefer that these lines all be horizontal, indicating no degradation. However, this is usually not the case, so I check to see that the lines are at least parallel, which indicates that any degradation has been relatively consistent from sample to sample.

After checking the data, I computed expression values. These data were then exported for further analysis using 'Significance Analysis of Microarrays' (SAM), a program that performs two-sample t-tests with multiplicity adjustment using false discovery rate (FDR)[2]. The basic idea behind FDR is to estimate how many of the 'significant' genes are in fact not differentially expressed. I have attempted to set the %FDR at 5%, meaning that approximately 5% of the genes found to be significant are false positives.

Output from SAM is presented in three separate Excel spreadsheets. Each spreadsheet has a name that indicates the comparison made. There are three worksheets in each workbook, one that contains the expression values, one has the SAM Plot, and the third, called SAM Output, contains the results. The results worksheet is relatively self-explanatory, but I will give some explanation here.

The result headings are given below:

970 Positive Significant Genes (Up in NG18-10)

Row	Gene Name	Gene ID	Score(d)	Numerator(r)	Denominator(s+s0)	Fold Change	q-value (%)	Filter	2

The data that is colored red indicates genes that were up-regulated in one of the samples. By default, this sample will be the second sample in the filename (e.g., if the file is SAM Analysis NG3-1 vs NG18-1, then red samples are upregulated in NG18-1 samples). Green data indicates down-regulation in the same sample. The meaning of the column headings is as follows:

Row – This corresponds to the row number in the expression values worksheet for this gene
Gene Name – This is the name given by Affymetrix
Gene ID – This is the accession number. It is also a hyperlink, so you can click on it to go to Stanford's website where there may be more information for this gene
Score(d) – This is the t-statistic. The larger this value, the more significant the gene.
Numerator(r) – This is the numerator of the t-statistic
Denominator(s + s0) – This is the denominator of the t-statistic
Fold change – This is the fold change (in the example above, it would be NG18-1/NG3-1)
q-value – This is the FDR equivalent of a p-value. It gives the percentage of the genes in the current row *and above* that are expected to be false positives.
Filter – This is a way to filter the data further, based on a required fold change. The current required fold change is two fold (the 2 in the adjacent cell indicates this). If you

want to adjust the filtering criterion, simply change the 2 to whatever fold change requirement you prefer. Note here that the fold change filter for the green data is for *down-regulated* genes. Therefore, the range for the number is [0,1]. Since there are so few samples, it is a very good idea to incorporate fold change in the analysis.

When you have the filter set to whatever criterion you like, you can then autofilter the data to select the genes that fulfill the criterion. If you are unfamiliar with autofiltering, please see the Excel help.

When I compared the two NG18 samples, there were quite a few genes that appeared to be differentially expressed. This indicates that the two isolates are different to some extent. My understanding is that a comparison of NG18 to NG3 is of interest, so I did this comparison two different ways. First, I found the genes that are in the intersection of the comparisons of NG3-1 vs NG18-1 and NG3-1 vs NG18-10 (with the additional constraint that the fold change had to be greater than 1.5 fold for the up-regulated genes, or less than 0.66667 fold for the down-regulated genes.). These results are saved in the file 'Common genes NG3 vs NG18.xls'. I also pooled the NG18 samples and performed a t-test comparing the two groups. This file is 'SAM Analysis NG3-1 vs NG18 pooled.xls'. The intersection results are probably a more reasonable list of differentially expressed genes, because the pooled data may be 'driven' by only one of the two NG18 cell samples.

I used the intersection results for an additional analysis, designed to find cellular functions that are being affected in the experiment. This is done using gene ontology (GO) terms for all of the genes that are called significant in an earlier analysis (in this case, the intersection results). We count the number of times a given GO term is found in the sample and compare to the total number of times this GO term is found on the HG-U133A chip. If the former is large in comparison to the latter, we can assume that there is something going on with that particular function.

As an example, we found seven genes that have a GO molecular function of 'tRNA ligase activity', and there are a total of 24 such genes represented on the U133A chip. The probability of getting this result if tRNA ligase activity is unaffected is extremely low (as evidenced by the p-value, which is 0.0000126).

As with the t-test results, we want to have some sort of adjustment for multiple comparisons. In this case it is difficult to come up with a FDR adjustment, so I did something slightly different. We have about 360 genes that were used for this analysis, and we want to know about how many 'significant' genes we should expect by chance alone. To do this, I randomly selected 360 genes from the population of genes on the U133A chip and calculated p-values as above. I repeated this process 100 times, and then calculated how many genes I got on average with a p-value less than 0.001 and 0.01. I got 0.17 genes on average with a p-value less than 0.001, and 3.29 with a p-value less than 0.01. This indicates that if we set the cutoff for 'significance' at 0.01, then we expect that about three of the genes are probably there by chance.

If you have any questions about the analysis, please contact me by email; jmacdon@med.umich.edu

---

[1] http://biosun01.biostat.jhsph.edu/~ririzarr/papers/affy1.pdf
[2] http://www-stat.stanford.edu/~tibs/SAM/pnassam.pdf